Item-Specific and Relational Encoding are Effective at

Reducing the Illusion of Competence

Nicholas P. Maxwell, Emily E. Cates, & Mark J. Huff

The University of Southern Mississippi

Word count: XXX

**Author Note**

Correspondence concerning this article can be addressed to Mark. J. Huff, School of Psychology, The University of Southern Mississippi, 118 College Dr. #5025, Hattiesburg, MS 39406, United States. Study materials and analyzed data are available via OSF (osf.io/k73r4). Supplemental Materials have been made available at osf.io/z7nm3/. This study was completed as part of the Honors Thesis requirements for EEC. Email: mark.huff@usm.edu

Abstract

Metamemory, or the ability to understand the capacities of one’s own memory, is important for learning. A common method for assessing metamemory is the Judgment of Learning (JOL) task in which participants are asked to judge the likelihood of correctly remembering a target word in a cue-target word pair when only presented with the cue at test. The associative direction of the cue-target pair can affect JOL accuracy. Unlike forward pairs (e.g., credit-card), in which JOLs accurately predict recall, an illusion of competence has been reported for backward associates (e.g., card-credit), symmetrical associates (e.g., salt-pepper), and unrelated pairs (e.g., artery-bronze) in which JOLs overestimate later recall. The present study evaluates whether the illusion of competence pattern can be reduced when participants use a deep item-specific or relational encoding strategy relative to reading. Overall, item-specific and relational encoding were found to reduce the illusion of competence for backward and unrelated pairs while improving the calibration between JOLs and recall. However, these encoding strategies largely reduced resolution, except for when pairs were unrelated. Thus, item-specific and relational encoding strategies are effective at reducing the illusion of competence by improving JOL calibration but not resolution.

Word count: 192

*Keywords:* Judgements of Learning; Illusion of Competence; Item-Specific Encoding; Relational Encoding; Calibration; Resolution

Item-Specific and Relational Encoding are Effective at

Reducing the Illusion of Competence

Successfully monitoring the progress of one’s learning at study is paramount for improving retention. Effective monitoring allows individuals to adjust their encoding strategies to maximize later retrieval (Nelson & Narens, 1990). Metamemory judgments, or having individuals judge or estimate the effectiveness their memorial abilities, can be used to obtain information about an individual’s knowledge of the learning process and how they adjust their encoding when faced with different study materials. A common method used to gauge metamemory knowledge is through the Judgment of Learning (JOL) task. In a standard JOL task, individuals study a set of cue-target word pairs and are asked to estimate the likelihood that they can recall a target word when only provided with the cue on a later test. These estimates can be elicited using several types of measurement scales such as Likert scales or binary “yes-no” responses (Hanczakowski, Zawadzka, Pasek, & Higham, 2013) or via a continuous 0 to 100 scale representing the percent likelihood that the target item will be successfully recalled at test (e.g., 100% = definitely would remember; 0% = definitely would not remember). The use of a 100-point scale is beneficial as it allows for a straightforward comparison between predicted target recall (via JOLs) and the percentage of targets that are correctly recalled at test.

Although JOL ratings accurately predict later recall (i.e., well-calibrated), several factors can affect the efficacy of JOLs. These include perception of identical cue-target word pairs as being fluent due to word repetitions (Castel, McCabe, & Roediger, 2007), increasing the time spent studying word pairs (Koriat & Ma’ayan, 2005), and the direction and strength of the relatedness between cue-target study pairs (Koriat & Bjork, 2005; Maxwell & Huff, 2021). The present study further examines factors that affect the accuracy of JOLs by examining the associative direction between cue-target pairs (i.e., probability that the cue word elicits the target at test or vice versa) and by testing whether encoding tasks that emphasize the shared or distinctive characteristics of the word pairs through relational and item-specific encoding tasks, respectively, can improve the accuracy of JOLs in predicting later recall.

**The Effect of Cue-Target Relations on JOL Accuracy**

Interest in the relationship between memory predictions and accuracy is not new. In an early example, Arbuckle and Cuddy (1969) asked participants to study letter-number pairs (e.g., A-73) and report whether they would or would not remember the pairs on a later test. At test, participants also provided a postdiction that they were either correct or incorrect regarding their answer. Arbuckle and Cuddy reported that participants correctly predicted later recall for an average of 67% of trials and correctly postdicted their responses for 88% of trials, leading the authors to conclude that participants had insight into how difficult each pair would be to remember and adjusted their predictions accordingly based on the association between participants’ predictions and subsequent recall.

More recently, Koriat and Bjork (2005) found that aspects of the associative relationship between cue-target study pairs, namely the direction and the strength of the relationship, can affect JOL accuracy. Specifically, the authors delineated between two types of associations thought to influence the relationship between JOLs and recall. First, *a priori* associations refer to associations in the forward direction (e.g., credit-card, stork-baby). The strength of these pair types is based on the likelihood that the cue word will elicit the target word at test. A priori/forward association strength can be readily assessed through the use of free association norms (e.g., The University of South Florida Free Association Norms; Nelson, McEvoy, & Schreiber, 2004; The Small World of Words Project; De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019). These norms are generated via free-association tasks in which participants are provided with a single cue word and asked to respond with the first target word that comes to mind. These norms can then be used to compute the probability of responding to word A with word B (i.e., forward associative strength, FAS). Separately, *a posteriori* associations refer to the perceived relatedness between pairs that are more apparent to participants when words are presented together. These pairs can refer to weakly associated pairs (e.g., article-newspaper) or strong associates in which the pair order has been flipped (i.e., backward pairs such as card-credit, baby-stork, etc.). Like a priori pairs, free association norms are useful for indexing backward associative strength (BAS) between pairs (i.e., the probability of responding to word B with word A in A-B pairs; see Nelson, McEvoy, & Dennis, 2000, for a review). Thus, a posteriori pairs could have either weak levels of FAS or strong levels of BAS.

To test the correspondence between JOLs and recall for a priori and a posteriori pairs, Koriat & Bjork (2005) evaluated JOL accuracy when participants studied unrelated and a priori study pairs (e.g., strong forward associates; Experiment 1), a priori and a posteriori pairs (e.g., backward associates; Experiment 2), and unrelated pairs, a priori pairs, and semantically related a posteriori pairs that shared no association based on the norms (Experiment 3). Across experiments, a posteriori pairs showed an *illusion of competence* pattern in which JOLs exceeded subsequent recall rates, indicating that participants overpredicted the likelihood that they would later recall the target word. This pattern was particularly robust for a posteriori backward pairs, as the cue word, when presented in isolation, does not ostensibly converge upon the studied target word. Thus, although participants predict that backward pairs are highly likely to be recalled, at test, recall accuracy is typically much lower than predicted.

The illusion of competence pattern found with a posteriori and backward pairs has similarly been reported by Castel et al. (2007) who examined the correspondence between JOLs and subsequent recall when participants studied and provided JOLs for strongly and weakly related forward associates, unrelated items, and identical cue-target word pairs. Overall, an illusion of competence emerged for identical word pairs in which JOLs exceeded subsequent recall rates. The authors ascribed this pattern to identical pairs being easier to learn, and therefore, more fluent versus forward and unrelated pairs given identical pairs were repeated items. As a result, participants may not have encoded identical pairs as deeply because they thought they would be easier to recall given the cue word was perfectly predictive of the target.

More recently, Maxwell and Huff (2021), further investigated the correspondence between JOLs and recall rates by looking at symmetrical associates (e.g., on-off), relative to forward, backward, and unrelated pairs. Symmetrical pairs differ from forward and backward pairs in that the associative strength between the cue and target word is equivalent in both directions (i.e., on-off would have approximately the same associative strength as off-on), whereas for forward and backward pairs the association is stronger is one direction than the other (i.e., tuna-fish is strongly associated in the forward direction, but has a weaker association in the backward direction, fish-tuna). Importantly, symmetrical pairs were matched to forward and backward pairs in overall associative strength. Thus, the only difference across pair types was the direction of association. Across four experiments, Maxwell and Huff found a robust illusion of competence pattern for backward pairs and, additionally, the illusion of competence was found on symmetrical associates, suggesting that the bidirectional association found for symmetrical pairs is not sufficient for the cue word to regularly illicit the target word. Maxwell and Huff also suggested that participants may be using both the forward and backward associations when studying symmetrical pairs even though only the forward association would be beneficial at test. Thus, the associative direction of a word pair can affect JOL accuracy, even when associative strength is matched across pair types.

Given that the illusion of competence can be found across several pair types, the goal of the present study was to examine methods that could potentially improve the JOL accuracy on subsequent recall to reduce the illusion of competence. One such method, tested in the present study, is by having participants engage in different types of encoding strategies that may help or hinder the processing of the relationship between the cue-target pair, a discussion to which we now turn.

**Item-Specific/Relational Processing on Memory Performance**

Memory researchers have long known that certain study tasks are more successful at improving retention than others. The levels-of-processing framework classifies tasks that promote elaborative processing of studied items that typically promote memory as “deep” tasks, while less successful tasks that focus on surface or perceptual features of study items are referred to as “shallow” tasks (Craik & Lockhart, 1972; Craik, 2002). Several deep tasks have been identified, including generation (Slamecka & Graf, 1978), production (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), and survival processing (Nairne, Thompson, & Pandeirada, 2007), however deep tasks can be bifurcated further based on a task’s propensity to encourage the processing of item-specific or relational features.

Regarding the effects of encoding depth on JOLs, few studies have assessed JOL accuracy across different LOP tasks. In a recent exception, Tekin and Roediger (in press) examined the reactive effects of JOLs using a levels-of-processing manipulation. They found that providing JOLs alongside a deep (vs. shallow) encoding task enhanced correct recognition, suggesting that deep encoding may be an effective method for reducing the illusion of competence by increasing memory performance (i.e., boosting memory performance to be more aligned with predicted memory). We further test this possibility within the context of cued-recall testing (vs recognition memory), while also investigating potential interactive effects between calibration and associative direction. Importantly, we further investigate the effects of deep processing on JOLs by delineating between two types of deep processing tasks that have been shown to benefit cued-recall performance: Item-specific and relational encoding.

According to the item-specific/relational processing framework (Einstein & Hunt, 1980; Hunt & Einstein, 1981), encoding tasks differ in the likelihood that they can encourage the processing of unique features of study items via item-specific processing, or through the processing of shared characteristics of study items via relational processing. Item-specific processing entails having participants focus on the unique features of items at study (e.g., for the pair cat-turtle, cats are mammals and turtles are reptiles, cats have fur and turtles have shells, etc.) while relational processing has participants focus on the shared features (e.g., cats and turtles are animals, both can be kept as pets, etc.). These types of processing qualitatively affect encoding strategies by changing how information encoded. Many studies have found differential memory benefits for item-specific and relational encoding tasks. For example, McCurdy, Sklenar, Frankenstein, and Leshikar (2020), showed that relational processing facilitated the generation effect for lower-constraint tasks (i.e., generating a target word in the presence of a cue) potentially because participants had to create a relationship between the two words. Relational processing could therefore be beneficial in studying unrelated word pairs since there is no existing relationship between the words. Separately, Huff and Bodner (2014) found that item-specific tasks were more likely to improve recall and recognition when study items were strongly related, but not when study items were weakly related. Similarly, relational tasks were more likely to improve recall and recognition when study items were weakly rather than strongly related (argued to be evidence of encoding variability of processing). Thus, although item-specific and relational processing tasks are generally classified as “deep” tasks based on the levels-of-processing framework, their relative memory benefits are affected by the association between study materials.

**Assessing JOL Accuracy**

Metacognitive research typically differentiates between two types of JOL accuracy. First, absolute accuracy or *calibration* describes the overall difference between predicted recall (assessed via JOLs) and actual performance at test. In terms of probabilities, calibration reflects the likelihood that a probabilistic prediction of an event will correctly map onto the event’s true probability of occurring (Jiang, Osl, Kim, & Ohno-Machadao, 2012). For example, a participant would be said to have perfect absolute accuracy if items given a JOL rating of 100 were recalled 100% of the time at test (i.e., their JOLs are well calibrated with recall). Item calibration has been a topic of extensive research across various domains of psychological research, including clinical psychology (Lindheim, Peterson, Mentch, & Youngstrom, 2020), eyewitness memory (Brewer & Wells, 2006), metacognitive confidence ratings (Double & Birney, 2017), and importantly, JOLs (Maxwell & Huff, 2021). Regarding JOLs, calibration can be easily assessed by plotting mean JOL ratings against mean recall proportions, so long as JOLs and recall are measured using the same scale. These *calibration plots* allow researchers to assess whether JOLs are over or underconfident (see Roediger, Wixted, & Desoto, 2012 for an example using confidence ratings), and furthermore, they can be used to assess whether metacognitive illusions like the illusion of competence uniformly affect recall at all JOL levels or whether the illusion is stronger for high JOL ratings vs low ratings (e.g., Maxwell & Huff, 2021).

Second, the relative accuracy between JOLs and recall has commonly been of interest to metacognitive researchers. Relative accuracy or *resolution* refers to the degree in which a person’s JOL discriminates between what is and what is not remembered (Rhodes, 2016) Unlike calibration, which can be assessed through plots, resolution is commonly assessed via Goodman-Kruskal gamma correlations. The gamma coefficient represents a measure of association between -1 and +1, with resolution decreasing as gamma approaches zero. Positive values denote the degree that remembered items were given high JOLs while non-remembered items received low JOLs, while negative gamma values denote the inverse of this pattern (Nelson, 1984).

Given that calibration and resolution reflect different elements of accuracy, they are influenced by different factors. Calibration is most strongly influenced by variables that affect the magnitude of JOLs and/or the likelihood that encoded information will be successfully recalled at test (Rhodes, 2016). Thus, factors that have been shown to directly influence the magnitude of JOLs like associative direction (Koriat & Bjork, 2005; Maxwell & Huff, 2021) and perceptual fluency (Rhodes & Castel, 2008) would be expected to produce changes in calibration. Similarly, encoding manipulations designed to affect recall (e.g., the item-specific/relational framework; Einstein & Hunt, 1980) would also be expected to influence calibration.

Whereas calibration is strongly influenced by factors that affect the magnitude of JOLs/recall, resolution is primarily impacted by factors influencing retrieval, including testing (King, Zechmeister, & Shaugnessy, 1980), practice (Koriat, Sheffer, & May’ayan, 2002), and timing (Nelson & Dunlosky, 1991; see Rhodes, 2016 for a comparison of factors influencing resolution). By allowing participants to complete test trials, engage in multiple study-tests cycles, or provide JOLs after a delay, resolution generally shows an improvement. Thus, resolution is expected to improve anytime the encoding task affords participants with an opportunity to adjust their JOL ratings based on previous performance.

Given the interactive benefits of item-specific and relational encoding with different associative materials, the present study tested whether these encoding strategies can improve both the calibration and resolution between JOLs and later recall, especially on backward and unrelated pairs in which the illusion of competence is robust (Castel et al., 2007; Koriat & Bjork, 2005; Maxwell & Huff, 2021). Specifically, we assessed JOLs and cued-recall performance for groups of participants who encode cue-target pairs using either item-specific or relational tasks relative to a standard JOL control task across forward, backward, symmetrical, and unrelated pair types. Additionally, we used calibration plots modeled after Maxwell & Huff (2021) to assess changes in calibration across each item type as a function of encoding strategy. Finally, changes in resolution were assessed using Goodman-Kruskal gamma correlations.

**Item-Specific Versus Relational Encoding Instructions**

The goals of the present study were twofold. First, we sought to replicate the illusion of competence pattern for backward, symmetrical, and unrelated pairs for participants completing a silent reading intentional encoding control task. Next, we tested whether item-specific/relational encoding tasks could reduce the illusion of competence by either lowering JOL ratings, increasing correct recall, or both. To date, little research has been conducted on JOLs within the context of the item-specific/relational framework. Recently, however, Senkova & Otani (2021) investigated whether JOL reactivity (i.e., increases in recall that result from making JOLs at encoding; see Soderstrom, Clark, Halamish, & Bjork, 2015 for an overview of JOL reactivity) reflected contributions of item-specific processing at encoding. While the focus of their research was primarily on reactivity rather than improving the calibration between JOLs and recall, we note that the authors compared recall rates for a JOL task across conditions designed to induce both item-specific and relational encoding. As noted by Senkova and Otani, list structure often induces item-specific and relational encoding strategies. Specifically, related lists often lead participants to engage in relational encoding, while unrelated lists are likely to induce item-specific strategies. Because memory is benefited most when participants use both item-specific and relational processes in tandem (e.g., Einstein & Hunt, 1980; Hunt & Einstein, 1981), list structure can moderate the efficacy of these encoding strategies. Item-specific encoding strategies are most beneficial when pairs are related, while unrelated pairs generally benefit most from relational encoding (Huff & Bodner, 2015; Hunt & Seta, 1984)

While Senkova and Otani (2021) did not have participants engage in additional item-specific/relational encoding tasks in addition to providing JOLs, we note that the authors manipulated list structure used it as a proxy to encourage these processing strategies. Specifically, participants studied lists of single words that were either categorized (i.e., related and thus expected to produce item-specific encoding) or uncategorized (i.e., unrelated and therefore expected to result in relational encoding). Overall, [JOL RESULTS] [RECALL RESULTS]

Given the reported benefits of item-specific and relational encoding on recall, we expected that having participants engage in these encoding tasks would reduce the illusion of competence by improving correct recall relative to the control group, without subsequently inflating JOLs. Because relational encoding encourages participants to generate associations between cue-target pairs, we expected that relational encoding would be beneficial across pairs given only the cue-word is available at test, but particularly beneficial for backward and unrelated pairs where the cue is less effective at prompting target retrieval. Finally, because item-specific (vs. relational) processing has been shown to be more beneficial to memory when pairs are strongly related (e.g., Huff & Bodner, 2014), it was expected that the item-specific task would be most beneficial for improving JOL calibration on related pairs. Thus, this encoding strategy was expected to reduce the illusion of competence for symmetrical and backward pairs, as forward pairs generally do not show an illusion of competence pattern (Maxwell & Huff, 2021). The qualitative differences in item-specific and relational encoding were expected to produce differential benefits on improving JOL accuracy depending on the pair type that was studied.

Finally, given that calibration is strongly affected by encoding manipulations designed to improve recall, we expected that any reductions in the illusion of competence resulting from participants’ use of item-specific and relational encoding strategies would subsequently result in improved calibration. However, gammas were not expected to differ as a function of encoding strategy, as resolution is primarily influenced by factors such as timing and practice, rather than changes in encoding strategy (Rhodes, 2016). Thus, we expected that any improvements in JOL accuracy would only be reflected in terms of absolute accuracy.

**Methods**

**Participants**

Eighty-eight University of Southern Mississippi undergraduates participated for partial course credit. Participants were randomly assigned to either the item-specific encoding group (*n* = 29), the relational encoding group (*n* = 31), or the read-only control group (*n* = 28). All participants were native English speakers with normal or corrected-to-normal vision. Sample sizes for each group were based on Maxwell & Huff (2021), and a sensitivity analysis conducted using *G\*Power* (Faul, Erdfelder, Lang, & Buchner, 2007) confirmed that our sample had sufficient power (.80) to detect a small-to-medium main effects and interactions (Cohen’s *d* = 0.27) or larger.

**Materials**

The stimuli used were 180 associative word pairs initially used by Maxwell and Huff (2021). Pairs were taken from the University of South Florida Free Association Norms (Nelson et al., 2004). These consisted of 40 forward pairs (e.g., credit-card), 40 backward pairs (e.g., card-credit), 40 symmetrical pairs (e.g., salt-pepper), 40 unrelated pairs (e.g. art-lion), and 20 weakly related, non-tested buffer pairs used to control for primacy and recency effects. Pairs were divided evenly into two study blocks, each containing 20 forward, backward, symmetrical, and unrelated pairs and 10 buffer pairs, for a total of 90 pairs in each list. All participants saw both lists presented in separate study-test blocks, the order of which was counterbalanced across participants. Each list began and ended with five buffer pairs, with the other pairs randomized anew for each participant.

Pair types were equated on associative strength (i.e., FAS and BAS) using the Nelson et al. (2004) free-association norms. Additionally, these pairs were designed to control for lexical and semantic properties that could potentially influence recall rates, including word length, SUBTLEX frequency (Brysbaert & New, 2009), and concreteness values derived from the English Lexicon Project (Balota et al., 2007; Maxwell & Huff, 2021). Further, both study blocks were matched on these properties. Thus, mean associative overlap and lexical/semantic properties were equivalent between direction types and study blocks. Finally, counterbalanced versions of the study lists were created that switched the order of the word pairs (i.e., forest-tree vs. tree-forest). As a result, forward pairs from one counterbalance became backward pairs on another and vice versa. Alternating pair direction allowed for greater control of item differences, particularly on forward and backward pairs, as the same items were used in the forward and backward directions across counterbalances. Pair order was similarly flipped and counterbalanced across unrelated and symmetrical pairs. Semantic and lexical characteristics of each pair type are reported in the Appendix (Tables A1-A2).

The cued-recall test in each block contained all 80 cue words from the studied pairs minus the buffer pairs which were not tested. The cue word was shown next to a question mark that had replaced the target word. Test order was newly randomized for each participant.

**Procedure**

The experimental procedure followed the general procedure used by Maxwell and Huff (2021). All participants completed the study individually on computers using *E-Prime* 3 software (Psychology Software Tools, Pittsburgh, PA). Participants were randomly assigned to one of three encoding groups: A read-only control group, an item-specific encoding group, or a relational encoding group. For each study group, participants were instructed that they would study a series of cue-target word pairs and that their memory for the target word in these pairs would be tested later with the cue word present. The cue word was always presented on the left and the target on the right. Participants were instructed to rate (via JOL) how likely they were to remember the target word if they were only presented with the cue at test. JOL ratings were made using a 0 to 100 scale, with 0 being “I am certain I WILL NOT REMEMBER the word pair” and 100 being “I am certain I WILL REMEMBER the word pair.” Participants were also instructed to use the full range of the scale to help reduce anchoring at points on the scale.

For the read group, participants were instructed to study the word pairs by reading them silently to themselves. For the relational group, participants were instructed to study the word pairs by thinking about how the pair of words were related to each other. Relational participants were also given the example of the word pair “Cat-Turtle”, and how they might think about how cats and turtles are both animals and can both be pets. For the item-specific group, participants were instructed to study the word pairs by thinking about how the words in each pair were unique with the example that for the pair “Cat-Turtle”, participants might think about how cats have fur, but turtles have shells and how cats are mammals, but turtles are reptiles. Item-specific and relational groups similarly completed encoding tasks silently. Participants only saw one type of task instruction. After the encoding instructions, participants completed a ten-word practice set using their assigned encoding task. Participants were then given their first block of word pairs to study at their own pace and provided their JOL ratings while the word pair was displayed. Finally, after studying half of the pairs, participants were presented a quick reminder to use their respective encoding strategy.

After the first study list was completed, participants were given two minutes to complete an arithmetic filler. Participants then completed a cued-recall test in which only the cue word was presented and were asked to provide the target word from memory. Participants were encouraged not to leave test answers blank and to try their best to retrieve the target word from memory. After the first cued-recall test was finished, participants completed a second study/test block using the same encoding instructions as the first. Once participants completed the second study/test block, they were debriefed and granted participation credit. Participants typically completed the experiment in under 1 hour.

**Results**

Prior to conducting analyses, study items that were missing JOL ratings or had ratings that were outside of the 0-100 range were removed. The screening processing removed fewer than 0.5% of items. When scoring recall responses, test items that were skipped were scored as incorrect and a liberal criterion for scoring correct items was adopted such that misspellings or pluralizations were scored as correct. All analyses were collapsed across block (analyses split by block are available in the Supplemental Materials; osf.io/z7nm3/), and we note that the data patterns were similar between blocks. Partial-eta squared (*η*p2) and Cohen’s *d* eﬀect sizes were included for signiﬁcant analyses of variance (ANOVAs) and *t*-tests, respectively. For all analyses, a *p* < .05 signiﬁcance level was used unless noted otherwise. For non-significant comparisons reported, we further analyzed the strength of the evidence supporting the null hypothesis using a Bayesian estimate (Masson, 2011; Wagenmakers, 2007). In this analysis, a model that assumes an effect is compared to a model that assumes a null effect. This process yields a probability estimate that the null hypothesis is retained (termed *p*BIC; Bayesian Information Criterion). The *p*BIC estimate is advantageous in that it is sensitive to sample size, increasing confidence in null effects reported. This Bayesian analysis is therefore supplementary to null effects detected with standard null-hypothesis-significance testing.

Mean JOL and recall rates as a function of pair type are reported in Figure 1. For completeness, all comparisons are reported in appendix Table A3. A 2 (Measure: JOL vs. Recall) × 3 (Encoding Group: Item-Specific vs. Relational vs. Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) mixed ANOVA compared differences between mean JOL ratings and recall rates across pair types and encoding groups. An effect of measure was found, *F*(1, 85) = 18.79, *MSE* = 694.46, *η*p2 = .18, such that collapsed across encoding groups and pair types, mean JOL ratings exceeded later recall rates (62.66 vs. 54.19). Next, an effect of encoding group was detected, *F*(2, 85) = 5.40, *MSE* = 814.98, *η*p2 = .11, in which JOL ratings/recall rates were significantly higher for the relational (61.44) and item-specific (60.12) groups relative to the read-only group (53.33). All comparisons differed significantly, *t*s ≥ 2.96, *d*s ≥ 0.78, except for the relational and item-specific groups, which were equivalent, *t* < 1, *p*BIC = .87. Finally, an effect of pair type was found, *F*(3, 255) = 766.58, *MSE* = 107.66, *η*p2 = 0.90, in which JOL ratings/recall rates were higher for symmetrical pairs (74.22), followed by forward pairs (72.29) backward pairs (59.01), and unrelated pairs (27.55). Comparisons across all pair types differed statistically, *t*s ≥ 2.69, *d*s ≥ 0.17.

A significant two-way interaction between measure and pair type confirmed that the illusion of competence replicated across encoding groups, *F*(3, 255) = 56.94, *MSE* = 87.42, *η*p2 = .40. Critically, however, a significant three-way interaction was also found, *F*(6, 255) = 15.56, *MSE* = 87.42, *η*p2 = .27, in which the magnitude of the illusion of competence differed as a function of encoding group.

Starting with backward pairs, reliable illusion of competence patterns were detected across encoding groups, though at different rates. In the read-control group, a robust illusion of competence was detected in which JOLs greatly exceeded later recall accuracy (68.62 vs. 37.78), *t*(27) = 9.44, *SEM* = 3.41, *d* = 2.19. For the item-specific group, JOLs also exceeded recall (69.55 vs 59.01), *t*(28) = 2.16, *SEM* = 5.12, *d* = 0.58, though at a lesser magnitude relative to the read condition. A similar pattern was observed in the relational group, where the JOLs exceeded recall, but again at a lower rate than the read group (71.55 vs 50.49), *t*(30) = 5.41, *SEM* = 4.05, *d* = 1.18.

Next, for forward pairs, an illusion of competence pattern was not found for any of the encoding groups, with JOLs matching later recall for both the read group (70.04 vs. 65.23), *t*(27) = 1.32, *SEM* = 3.42, *p* = .19, *p*BIC = .69), and the relational group (72.96 vs 77.22), *t*(30) = 1.15, *SEM* = 3.86, *p* = .26, *p*BIC = .74. For the item-specific group, however, JOLs were lower than later recall rates (68.67 vs. 78.84), *t*(28) = 2.42, *SEM* = 4.41, *d* = 0.65—a situation in which JOLs can underestimate later recall.

For symmetrical pairs, the illusion of competence was moderated by encoding. For the read group, JOLs exceeded later recall accuracy (80.22 vs. 64.85), *t*(27) = 3.59, *SEM* = 4.48, *d* = 1.06; however, for both the item-specific and relational groups, the illusion of competence did not emerge as JOLs were equivalent to subsequent recall rates (71.62 vs 78.24), *t*(28)= 1.41, *SEM* = 4.90, *p* = .17, *p*BIC = .66, and (75.77 vs 74.41), *t* < 1, *SEM* = 3.46, *p* = .67, *p*BIC = .83, respectively.

Finally, for unrelated pairs, the illusion of competence was observed in both the read group (24.78 vs 14.73), *t*(27) = 3.23, *SEM* = 3.26, *d* = 0.76 and the item-specific group (40.64 vs 14.35), *t*(28) = 5.71, *SEM* = 4.81, *d* = 1.56, as JOLs exceeded later recall. However, the illusion of competence was not found in the relational group (36.59 vs. 32.52), *t* < 1, *SEM* = 4.52, *p* = .35, *p*BIC = .78), indicating that relational encoding provides a unique benefit on unrelated pairs by improving the correspondence between JOLs and subsequent recall.

Taken together, item-specific and relational processing tasks were both found to reduce and/or eliminate the illusion of competence pattern, but these reductions depended upon the pair type studied. Item-specific encoding was most successful at reducing the illusion of competence when participants studied backward associates. Relational encoding, however, was most beneficial for reducing the illusion of competence for unrelated pairs. The improved correspondence between JOLs and recall for item-specific and relational tasks was likely due to both tasks increasing correct recall (vs. adjusting JOL ratings) relative to reading, given both tasks are classified as deep processing tasks. Indeed, overall JOL rates across the three encoding groups were stable, *F*(2, 85) < 1, *MSE* = 147.50, *p* = .59, *p*BIC = .98, though recall rates were greater in the item-specific (*M* = 57.62) and relational groups (*M* = 58.67), relative to the read group (*M* = 45.68; *t*s ≥ 3.13, *d*s ≥ 0.57), with the item-specific and relational groups being equivalent, *t* < 1, *p*BIC = .88.

**Calibration**

We next assessed the absolute accuracy between JOLs and recall for each pair types using a series of calibration plots (cf. Maxwell and Huff, 2021). To generate these plots, JOLs were first rounded to the nearest 10% increment, which resulted in 11 JOL bins ranging from 0 to 100%. For example, the 0% JOL increment contains the proportion of correct recall for items given an initial judgment of 0%, the 10% increment contains the proportion of correct recall for items given an initial judgment of 10%, etc. Mean correct recall for each JOL bin was then plotted.

Figures 2-4 display calibration plots for each encoding group as a function of pair direction. Plots are structured such that they include a calibration line denoting a perfect one-to-one correspondence between JOL ratings and mean correct recall (e.g., a 40% JOL and a 40% correct recall rate would be perfectly calibrated). Overestimations are reflected by data points falling below the calibration line, while underestimations are represented by data points falling above the calibration lines. Calibration plots were analyzed using a 3 (Encoding Group: Item-Specific vs. Relational vs Read) × 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 11 (JOL increment) mixed ANOVA, however, the 3-way interaction was non-significant, *F*(60, 2520) = .81, *MSE* = 919.81, *p* = .86, *pBIC* = 1. Therefore, we next analyzed calibration plots separately for each of the three encoding groups.

Starting with the read group, for unrelated pairs, JOLs were found to overestimate later recall at all JOL increments (JOLs > 30%). However, for associative pairs overestimations emerged at higher JOL ratings. For backward pairs, overestimations occurred at JOLs greater than 50%, while overestimations of symmetrical and forward associates each occurred at the highest JOL ratings (< 90%). Using a 4 (Pair Type: Forward vs. Backward vs. Symmetrical vs. Unrelated) × 11 (JOL increment) mixed ANOVA, these patterns were confirmed by effects of Pair Type, *F*(3, 81) = 32.19, *MSE* = 50758.57, *η*p2= .51, JOL Increment, *F*(10, 270) = 9.74, *MSE* = 14084.99, *η*p2 = .27, and a significant interaction, *F*(30, 810) = 2.50, *MSE* = 2084.56, *η*p2 = .09.

Next, for the item-specific group, overestimations of unrelated pairs were observed for JOL ratings above 40%. For backward pairs, calibration of JOLs and recall was improved relative to silent reading, as overestimations occurred at JOL ratings greater than 80%. Finally, for symmetrical and forward associates, overestimation again occurred only for JOLs greater than 90%. These patterns were again confirmed by effects of Pair Type, *F*(3, 84) = 36.92, *MSE* = 57849.302, *η*p2= .57, JOL Increment, *F*(10, 280) = 8.00, *MSE* = 16024.10, *η*p2 = .22, and a significant interaction, *F*(30, 840) = 3.37, *MSE* = 2932.80, *η*p2 = .11.

Finally, for the relational group, JOL overestimations of unrelated pairs were reduced relative to the read and item-specific groups, as overestimations emerged JOL ratings above 50%. However, overestimations of associative pairs followed similar patterns as observed for the item-specific and read groups. Specifically, overestimations of backward pairs emerged at JOLs ratings greater than 60%, while overestimations of symmetrical and forward associates again occurred at JOLs greater than 90%. These patterns were confirmed by effects of Pair Type, *F*(3, 87) = 23.86, *MSE* = 31563.43, *η*p2= .45, JOL Increment, *F*(10, 290) = 10.14, *MSE* = 19751.25, *η*p2 = .26, and a significant interaction, *F*(30, 870) = 2.73, *MSE* = 2894.75, *η*p2 = .09.

Collectively, the calibration plots reveal important qualitative differences regarding specific JOL increments in which item-specific and relational encoding tasks start to reduce the illusion of competence pattern. For forward and symmetrical pairs, where illusions of competence are generally not found, all encoding groups showed similar calibration patterns. However, for unrelated and backward pairs, the illusion of competence pattern emerged at higher JOL increments relative to the read group. In particular, item-specific encoding was most effective at increasing the JOL increment in which the illusion of competence pattern was detected for backward pairs (> 80%), while relational encoding was most effective at increasing the JOL increment for unrelated pairs (> 50%), again demonstrating the differential benefits of item-specific and relational encoding at improving JOL accuracy.

**Resolution**

Finally, we assessed whether item-specific or relational encoding instructions affected the resolution between JOLs and recall. Following the procedure used by Dunlosky and Nelson (1992; 1994), we computed Goodman-Kruskal gamma correlations (*G*) between JOLs and recall for each participant for each of the four pair types (forward, backward, symmetrical, and unrelated; see Table 1 for mean gammas and 95% *CI*s). Starting with forward pairs, relative to silent reading, both item specific and relational encoding resulted in reduced resolution compared to silent reading (.10 vs. .13 vs. .35, respectively). This pattern subsequently extended to backward associates (.12 vs. .07 vs. .24) and symmetrical associates (.15 vs. .13 vs. .23). However, for unrelated pairs, resolution was increased when participants engaged in item-specific (.26) and relational encoding (.33) relative to participants in the read group (.20). Thus, while item-specific and relational encoding strategies are effective at reducing the illusion of competence, this reduction appears to occur primarily due to increased calibration rather than resolution.

**Discussion**

The purpose of the present study was to improve the predictive efficacy of JOL ratings on subsequent recall of forward, symmetrical, backward, and unrelated cue-target word pairs. Previous research has consistently found that JOLs tend to be over predictive on unrelated and deceptive backward pairs resulting in an illusion of competence pattern (Koriat & Bjork, 2005; Maxwell & Huff, 2021). Given previous work has shown benefits of deep processing on JOL calibration (Tekin & Roediger, in press), we attempted to further qualify the effects of deep processing by comparing item-specific and relational encoding tasks to a read-control group. Finally, we included calibration plots and gammas as measures of JOL calibration and resolution.

Overall, forward pairs did not produce an illusion of competence pattern across any of the three encoding groups. Furthermore, JOLs in the item-specific group underpredicted later recall of forward pairs (cf. Koriat & Bjork, 2005; Castel et al. 2007). Next, symmetrical pairs showed a small illusion of competence that was eliminated when participants studied pairs using item-specific and relational encoding. As expected, the illusion of competence was robust for backward and unrelated pairs, and furthermore, both the item-specific and relational tasks successfully reduced, but did not eliminate, this metacognitive illusion. Specifically, for backward pairs, both the item-specific and relational tasks were found to reduce the illusion of competence, though the item-specific task produced the greater reduction. These results were consistent with our predictions that item-specific encoding would be most beneficial in reducing the illusion of competence for related pairs (cf. Huff & Bodner, 2014). In contrast, however, the relational group produced a greater reduction for unrelated pairs relative to the item-specific group. Collectively then, both item-specific and relational encoding tasks can improve JOL accuracy versus a read task, though their relative effectiveness depends upon the associative direction of the pair type.

Following the design of Maxwell & Huff (2021), calibration plots were computed in order to further assess the absolute accuracy between JOLs and recall for each pair direction split by encoding task. Across all groups, participants were generally well calibrated for the forward and symmetrical pair types. For the read group, participants were overconfident for unrelated pairs at all JOL increments and for backward pairs above all JOL increments over half. Thus, overestimation was most likely to occur for pairs in which relatedness cues used at encoding were not readily available at retrieval. For the item-specific group, participants were overconfident for unrelated pairs at almost all of the JOL increments and overconfident for backward pairs at high JOL increments. Thus, relative to the read group, item-specific encoding produced a slight improvement in calibration for unrelated pairs. Finally, for the relational group, participants were overconfident for backward and unrelated pairs at all JOL increments over 50%. Thus, compared to silent reading and item-specific encoding, relational encoding greatly improved participants’ abilities to accurately predict their own recall for unrelated pairs. This suggests that there is a benefit to studying word pairs using a relational encoding strategy, particular when study pairs are unrelated.

Regarding resolution, both item-specific and relational encoding strategies largely reduced to relative accuracy compared to silent reading. For forward, backward, and symmetrical paired associates, mean *G* was lower when participants engaged in these encoding strategies relative to participants in the read group. For unrelated pairs, however, item-specific and relational encoding strategies increased resolution. Thus, it appears that while item-specific and relational strategies can increase the accuracy of JOLs, the effect is strongly moderated pair relatedness.

Of particular interest is the disconnect between calibration and resolution. While the item-specific and relational encoding strategies were each effective at reducing the illusion of competence and improving calibration for both related and unrelated pairs, this accuracy improvement was only reflected in resolution when pairs where unrelated. For related pairs, resolution decreased relative to the read group. The observation that item-specific and relational processing primarily benefited calibration may result from how these encoding manipulations reduced the illusion of competence. For example, when pairs were related, both item-specific and relational encoding reduced the illusion of confidence by increasing participants cued-recall performance, rather than acting as a control process that allowed participants to modify their JOLs. However, for unrelated pairs, item-specific processing increased JOLs without affecting recall, and relational encoding boosted both JOLs and recall. [EXPAND]

[NEW PARAGRAPH TYING BACK TO OTANI PAPER]

Finally, while it is evident that both item-specific and relational encoding tasks can be beneficial for improving JOL accuracy and reducing the illusion of competence, neither of these encoding tasks were able to eliminate the illusion of competence completely. [LIMITATIONS AND FUTURE DIRECTIONS]

**Conclusion**

In sum, the present study found that the illusion of competence can be reduced when applying the item-specific/relational encoding framework. Specifically, we showed that the illusion of competence for backward and symmetrical associates can be reduced via item-specific encoding and that overestimation of unrelated pairs is reduced when participants use a relational encoding strategy. Additionally, calibration plots revealed that these encoding manipulations increased the correspondence between JOLs and recall. Gamma correlations, however, indicated that [EXPAND]. Taken together, [SUMMARIZE]

**Open Practices Statement**

The data for all experiments have been made available at http://osf.io/k73r4 and none of the experiments were preregistered.

**Compliance with Ethical Standards:**

The studies reported were approved by the University of Southern Mississippi Institutional Review Board (Protocol #IRB-18-15) and found to be in accordance with the 1964 Helsinki Declaration ethical principles. Informed consent was obtained from all individuals who participated in this study. The authors report no competing interests.

**References**

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126–131.

doi: 10.1037/h0027455

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B, & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445-459. doi: 10.3758/BF03193014

Brewer, N. & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*(1), 11-30.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*(1), 107-111. doi: 10.3758/BF03194036

Craik, F. I. M. (2002). Levels of processing: Past, present … and future? *Memory, 10*(5-6). 305-318. doi: 10.1080/09658210244000135

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. doi: 10.1016/S0022-5371(72)80001-X

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods,* *51*(3), 987-1006. doi: 10.3758/s13428-018-1115-7

Double, K. S., & Birney, D. P. (2017). Are you sure about that? Eliciting confidence ratings may influence performance on Raven’s progressive matrices. *Thinking & Reasoning, 23*(2), 190-206.

Dunlosky, J. & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*(4), 374-380.

Dunlosky, J. & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*, 545-565.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 588-598. doi: 10.1037/0278-7393.6.5.588

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior* *Research Methods*, *39*(2), 175-191. doi: 10.3758/BF03193146

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language, 69*(3), 429–444. doi: 10.1016/j.jml.2013.05.003

Huff, M. J., & Bodner, G. E. (2014). All varieties of encoding variability are not created equal: Separating variable processing from variable tasks. *Journal of Memory and Language, 73*, 43-58. doi: 10.1016/j.jml.2014.02.004

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*(5), 497-514. doi: 10.1016/S0022-5371(81)90138-9

Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(3), 454-464.

Jiang, X., Osl, M., Kim, J., & Ohno-Machado, L. (2012). Calibrating predictive model estimates to support personalized medicine. *Journal of the American Medical Informatics Association, 19*, 263-274.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93*, 329-343.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 187–194. [doi: 10.1037/0278-7393.31.2.187](https://doi.org/10.1037/0278-7393.31.2.187)

Koriat, A., & Ma’Ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language,* *52*(4), 478-492. doi: 10.1016/j.jml.2005.01.001

Koriat, A., Sheffer, L., & Ma’Ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*, 147-162.

Lindhiem, O., Peterson, I. T., Mentch, L. K., & Youngstrom, E. A. (2020). The importance of calibration in clinical psychology. *Assessment, 27*(4), 840-854.

Macleod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *36*(3), 671-685. doi: 10.1037/a0018785

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods,* *43*(3), 679-690. doi: 10.3758/s13428-010-0049-5

McCurdy, M. P., Sklenar, A.M., Frankenstein, A. N., & Leshikar, E. D. (2020). Fewer Generation Constraints Increase the Generation Effect for Item and Source Memory through Enhanced Relational Processing. *Memory,* *28*(5), 598-616. doi: 10.1080/09658211.2020.1749283

Maxwell, N. P., & Huff, M. J. (2021). The deceptive nature of associative word pairs: Effects of associative direction on judgments of learning. *Psychological Research, 85*(4), 1757-1775. doi: 10.1007/s00426-020-01342-z

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *33*(2), 263-273. doi: 10.1037/0278-7393.33.2.263

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition,* *28*(6), 887-899. doi: 10.3758/bf03209337

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407. doi: 10.3758/bf03195588

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychonomic Bulletin, 95*(1), 109-133.

Nelson, T. O., & Dunlosky, J. (1991). When people’s judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “delayed-JOL effect.” *Psychological Science, 2*, 267-270.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation,* *26*, 125-173. doi: 10.1016/s0079-7421(08)60053-5

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://www.pstnet.com

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.) *The Oxford Handbook of Metamemory* (pp. 90-117). Oxford, Oxford University Press.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions: *Journal of Experimental Psychology: General, 137*(4), 615-625.

Roediger, H. L., Wixted, J. H., & DeSoto, K. A. (2012). The curious complexity between confidence and accuracy in reports from memory. In L. Nadel & W. P. Sinnott-Armstrong (Eds.), *Memory and law* (pp.84-108). Oxford, Oxford University Press.

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 553-558.

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition, 49,* 955-967.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory,* *4*(6), 592-604. doi: 10.1037/0278-7393.4.6.592

Tekin, E. & Roediger, H. L. (in press). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift für Psychologie, 228*(4), 278-290. doi: 10.1027/2151-2604/a000425

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review,* *14*(5), 779-804. doi: 10.3758/bf03194105

Table 1.

*Mean (± 95% CI) Goodman-Kruskal Gamma Correlations Between JOLs and Recall for each Encoding Group as a Function of Pair Type*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Encoding Group | Forward | Backward | Symmetrical | Unrelated |
| Read | .35 (.12) | .24 (.13) | .23 (.10) | .20 (.18) |
| Item-Specific | .10 (.11) | .12 (.13) | .15 (.16) | .26 (.16) |
| Relational | .13 (.11) | .07 (.07) | .13 (.14) | .33 (.10) |



**Mean % JOL/Recall**

**Mean % JOL/Recall**



**Mean % JOL/Recall**

**Pair Type**

*Figure 1*. Mean JOL and recall rates as a function of pair type in the Read group (top panel), Item-Specific group (middle panel), and the Relational group (bottom panel). Bars represent 95% confidence intervals.

Chart

Description automatically generated

*Figure 2.* Calibration plots as a function of pair direction in the Read Group. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued recall. Overconfidence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% confidence interval.

Diagram, engineering drawing

Description automatically generated with medium confidence

*Figure 3.* Calibration plots as a function of pair direction in the Item-Specific Group. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued recall. Overconfidence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% confidence interval.

Chart

Description automatically generated

*Figure 4.* Calibration plots as a function of pair direction in the Relational Group. Dashed lines indicate perfect calibration between JOL ratings and proportion of correct cued recall. Overconfidence is represented by points falling below the calibration line. Data were smoothed over three adjacent JOL ratings. Bars represent 95% confidence interval.

**Appendix**

Table A1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Condition | Variable | *M* | *SD* | *Min.* | *Max.* |
| Forward | FAS | .37 | .21 | .05 | .81 |
|  | BAS | .00 | .00 | .00 | .00 |
| Backward | FAS | .00 | .00 | .00 | .00 |
|  | BAS | .37 | .21 | .05 | .81 |
| Symmetrical | FAS | .19 | .13 | .01 | .46 |
|  | BAS | .19 | .13 | .02 | .52 |

*Mean Associative Strength Summary Statistics Forward, Backward, and Symmetrical Pairs.*

*Note.* FAS (forward associative strength) and BAS (backward associative strength) values for unrelated pairs as these items share zero associative overlap.

Table A2

*Summary Statistics for Cue and Target Concreteness, Length, and Frequency Item Properties as a Function of Pair Type.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pair Type | Position | Variable | *M* | *SD* |
| Forward | Cue | Concreteness | 4.97 | 1.22 |
|  |  | Length | 6.20 | 1.86 |
|  |  | Frequency | 3.74 | 0.67 |
|  | Target | Concreteness | 4.96 | 1.14 |
|  |  | Length | 4.46 | 1.27 |
|  |  | Frequency | 2.49 | 0.63 |
| Backward | Cue | Concreteness | 4.96 | 1.14 |
|  |  | Length | 4.46 | 1.27 |
|  |  | Frequency | 2.49 | 0.63 |
|  | Target | Concreteness | 4.97 | 1.22 |
|  |  | Length | 6.20 | 1.86 |
|  |  | Frequency | 3.74 | 0.67 |
| Symmetrical | Cue/Target | Concreteness | 4.70 | 1.38 |
|  |  | Length | 5.21 | 1.94 |
|  |  | Frequency | 3.23 | 0.67 |
| Unrelated | Cue/Target | Concreteness | 4.63 | 128 |
|  |  | Length | 5.21 | 1.52 |
|  |  | Frequency | 2.49 | 0.85 |

*Notes.* Frequency is measured using SUBTLEX word frequency measure (Brysbaert & New, 2009). Concreteness and length were taken from the English Lexicon Project (Balota et al., 2007).

Table A3

*Comparison of mean JOL ratings and correct recall percentages across all associative direction groups for each encoding group.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Encoding Group | Task | Group | *M* | | *95% CI* | F | B | S |
| Item-Specific | JOL | Forward | 68.67 | | 5.95 |  |  |  |
|  |  | Backward | 69.55 | | 6.41 | 0.05 |  |  |
|  |  | Symmetrical | 71.62 | | 5.24 | 0.19 | 0.13 |  |
|  |  | Unrelated | 40.64 | | 7.49 | 1.51\* | 1.51\* | 1.74\* |
|  | Recall | Forward | 78.84 | | 5.47 |  |  |  |
|  |  | Backward | 59.01 | | 6.85 | 1.16\* |  |  |
|  |  | Symmetrical | 78.24 | 6.05 | | 0.04 | 1.08\* |  |
|  |  | Unrelated | 14.35 | 4.35 | | 4.75\* | 2.83\* | 4.42\* |
| Relational | JOL | Forward | 72.96 | | 4.86 |  |  |  |
|  |  | Backward | 71.55 | | 5.52 | 0.08 |  |  |
|  |  | Symmetrical | 75.77 | | 4.82 | 0.20\* | 0.29\* |  |
|  |  | Unrelated | 36.59 | | 5.90 | 2.37\* | 2.15\* | 2.66\* |
|  | Recall | Forward | 77.22 | | 6.09 |  |  |  |
|  |  | Backward | 50.49 | | 6.96 | 1.44\* |  |  |
|  |  | Symmetrical | 74.41 | 5.94 | | 0.16 | 1.30\* |  |
|  |  | Unrelated | 32.52 | 8.08 | | 2.07\* | 0.71\* | 1.95\* |
| Read | JOL | Forward | 70.04 | | 3.89 |  |  |  |
|  |  | Backward | 68.62 | | 4.39 | 0.13 |  |  |
|  |  | Symmetrical | 80.22 | | 4.20 | 0.93\* | 1.00\* |  |
|  |  | Unrelated | 24.85 | | 5.68 | 3.44\* | 3.19\* | 4.11\* |
|  | Recall | Forward | 62.23 | | 6.96 |  |  |  |
|  |  | Backward | 37.78 | | 5.91 | 1.40\* |  |  |
|  |  | Symmetrical | 64.85 | 6.34 | | 0.15 | 1.64\* |  |
|  |  | Unrelated | 14.76 | 3.96 | | 3.11\* | 1.69\* | 3.51\* |

*Note.* Mean JOL and recall rates for each associative direction condition across each encoding group. The three right-most columns indicate Cohen’s *d* effect sizes for post-hoc comparisons, \* = *p* < .05.